

Safe In-vehicle Dialogue Using Learned Predictions of User Utterances

Staffan Larsson
Talkamatic AB
Första Långgatan 18
413 28 Göteborg
Sweden
staffan@talkamatic.se

Fredrik Kronlid
Talkamatic AB
Första Långgatan 18
413 28 Göteborg
Sweden
fredrik@talkamatic.se

Pontus Wärnestål
Halmstad University
Box 823
301 18 Halmstad
Sweden
pontus.warrestal@hh.se

Abstract

We present a multimodal in-vehicle dialogue system which uses learned predictions of user answers to enable shorter, more efficient, and thus safer natural language dialogues.

1 Background

1.1 Driver Distraction

Driver distraction is a common cause of accidents, and is often caused by the driver interacting with technologies such as mobile phones, media players or navigation systems. A study, commonly referred to as the “100 car study” (Neale et al., 2005) revealed that secondary task distraction is the largest cause of driver inattention, and that the handling of wireless devices is the most common secondary task.

As interaction complexity in the car increases due to more advanced infotainment systems and smartphones, drivers are often executing several tasks in parallel to the primary task of driving. The increased functionality of these systems has resulted in large hierarchical information architectures that prolong interaction time, thereby negatively affecting safety as well as user experience (Kern and Schmidt, 2009).

1.2 Relation to state of the art

State-of-the-art infotainment systems typically do not include user models at all. Siri, available on the Apple iPhone 4S and later models, has a static user model containing personal information explicitly provided by the user (home address, etc.). This information is used in voice interactions; for example, given that the user has entered their family relations, phrases like “Call my wife” can be used. A different approach is taken in Google Now, which dynamically learns user patterns from

observations and presents unrequested information as “cards” on the screen. However, Google Now does not attempt to integrate predictions into dialogue interaction.

The work reported here explores the use of adaptive user modeling in multimodal dialogue systems. User preferences and behaviour patterns are learnt from observations of user interactions with the infotainment system and the context in which these interactions take place, and are used proactively to predict user answers and thereby enable shorter and more efficient interaction. The underlying motivating assumption is that using apps and services in an in-vehicle context inherently leads to distraction, and that reducing interaction time will reduce driver distraction.

1.3 TDM

Based on Larsson (2002) and later work, Talkamatic AB has developed the Talkamatic Dialogue Manager (TDM).

TDM provides a general interaction model based on interaction which are basic to human-human linguistic interaction, resulting in a high degree of naturalness and flexibility which increases usability. The model is domain-independent which means that dialogue behaviour can be altered without touching application properties and vice versa. TDM also offers integrated multi-modality which allows user to freely switch between modalities (Larsson et al., 2011).

1.4 Grounding in TDM

Grounding (Clark and Brennan, 1990) is, roughly, the process of making sure that dialogue participants agree on what has been said so far and what it meant. TDM has an extensive model of grounding (Larsson, 2002). It operates on different levels:

- *Perception*
- *Semantic Understanding*

- *Pragmatic Understanding*
- *Acceptance*

System feedback (positive, negative and in some cases interrogative) can be generated on each level:

- Examples: “I didn’t hear” – negative perception
- “To work, is that right?” – interrogative semantic understanding
- “OK” – positive acceptance.

2 Learning and Classification

Many dialogue applications require the user to answer a number of questions. To make dialogue shorter, we have extended TDM so that it tries to predict user answers on the basis of a user model learned from observations of user behaviour. As an illustration, we use a road information application which tries to predict the user’s destination and thereby eliminate the need to ask the user about this.

2.1 Learning Method

Initially, a range of learning methods requiring (N-gram, MDP, POMDP) were explored and evaluated, but the KNN (K-Nearest Neighbours) (Mitchell, 1997) was considered the best method. An important advantage is that KNN can learn from a relatively small set of observations. This is in contrast to the MDP and POMDP (and to a lesser extent, N-gram) methods, which require large amounts of data to generate useful behaviour. A potential drawback of KNN is that this model cannot model sequences of user behaviours.

2.2 Parameter Selection

On the basis of user studies provided from the user partner of the project, it was decided that the most important user model parameters was position, day of the week and hour of the day. The training data were simulated and correspond to the behaviour of an archetypal persona provided by the user partner in the project.

2.3 Learning and Classification

The learning part of the system listens for a number of events, such as “start-car”, “stop-car” etc.. From these events and information about current position, the time of the day and the day of

the week, the system creates new data instances. The system thus learns how the user’s destination varies depending on these parameters. A sample dataset is shown in Figure 1, where data points show destinations of trips initiated at various times of the week.

When the dialogue manager requests a prediction of the destination, the KNN algorithm tries to find the K data points closest to the present data point, and the top alternatives are returned to the dialogue manager together with confidence scores indicating the reliability of the predictions.

3 Integration of Classifications into TDM

3.1 Grounding uncertain information

We treat the information emanating from the user model as uncertain information about a (predicted) user utterance. Hence, the same mechanisms used for grounding utterances have been adapted for integrating user model data.

3.2 Integrating Classifier Output

TDM is based on the Information State Update (ISU) approach to dialogue management. The information state in TDM is based on that of the system described in Larsson (2002) and includes Questions Under Discussion, a dialogue plan, and shared commitments.

The rule for integrating the user model data is a standard ISU rule, consisting of preconditions and effects on the information state. We describe these informally below:

PRECONDITIONS

- If there is a propositional answer from the user model resolving a question in the current plan...
- and if the confidence score reported from the user model is sufficient, then...

EFFECTS

- accept the propositional answer (include it into the shared commitments), and...
- give appropriate feedback to the user depending on the confidence score:
- High confidence \Rightarrow embedded feedback
 - “Which route do you want to take to work?”.

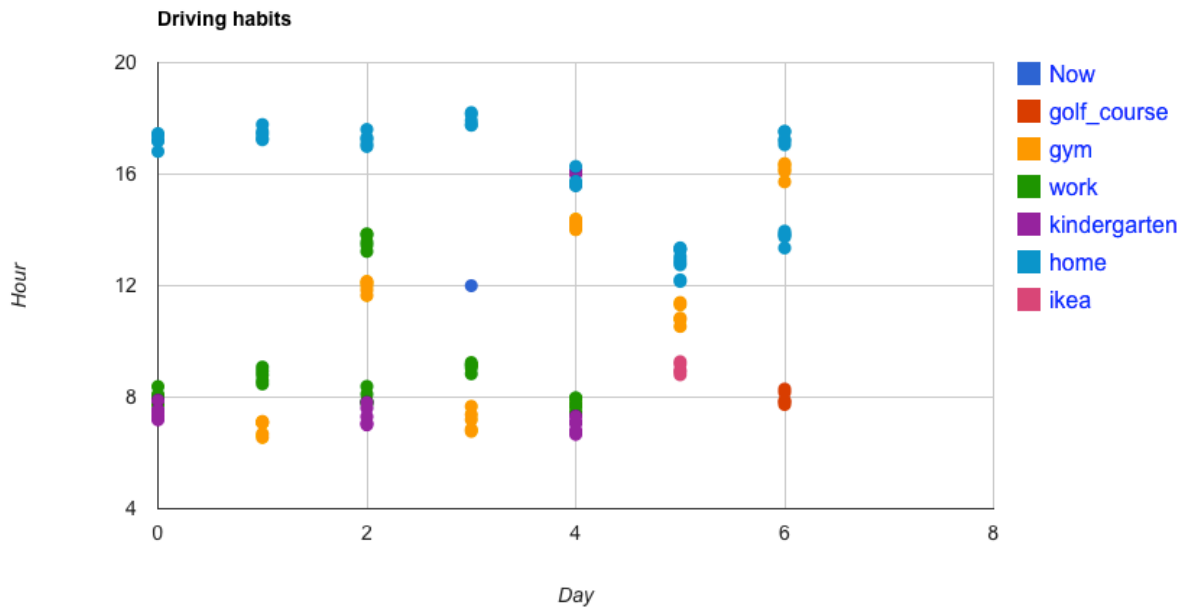


Figure 1: A sample dataset. The horizontal axis shows days of the week (0=Monday, ..., 6=Sunday) and the vertical axis shows hour of the day. Data points show destinations of trips initiated at the time indicated by their position. (“Now” is the current time, in this case Thursday at lunchtime.)

- The user can always reject the prediction by requesting another destination.
- Medium confidence \Rightarrow positive feedback
 - “I assume you’re going to work”.
 - If the user says “no”, the answer is rejected
 - Silence is interpreted as acceptance.
- Low confidence \Rightarrow interrogative feedback
 - “To work, is that correct?”
 - In this case, the user needs to explicitly accept the proposed answer.
 - Otherwise, the user is prompted for an answer.

3.3 GUI output

If the ISU rule above does not apply because of too low confidence scores, user model information is still used in the GUI. When a Wh-question is raised by the system, the GUI always presents a list of possible alternatives. High-confidence alternatives are highlighted and sorted before the other alternatives in the list.

4 Resulting behaviour

The demonstrator enables interaction with a learning dialogue system which uses predictions to sim-

plify interactions. Here is an sample interaction:

```
User: Traffic information
Car: Ok. What road?
User: E6.
Car: Showing traffic on the E6
```

If this is repeated on a number of occasions, eventually the system will use a prediction:

```
User: Traffic information
Car: Showing traffic on the E6
```

The system thus reduces the need for repetitive and information-scarce utterances from the user. As soon as the system has started identifying a pattern, it will start to suggest the most probable alternatives. Initially, the most probable answers are presented to the user as the top items in a list. The alternatives are also marked in a different color to make them more visible to the user (not shown here).

```
User: Traffic information
Car: Ok. What road?
Car GUI: [E6] [E45] [E20] [155]
User: E6.
Car: Showing traffic on the E6
```

After some further use, the system has identified a pattern which is prominent enough for the system to make a suggestion:

```
User: Traffic information
Car: E6, is that right?
User: Yes.
Car: Showing traffic on the E6
```

After getting further support for its hypothesis, the system will merely inform the user that an assumption has been made. If the user is satisfied with the assumption, she does not need to do anything, but can correct or confirm it if desired.

```
User: Traffic information
Car: I assume E6.
User: [silence]
Car: Showing traffic on the E6
```

```
User: Traffic information
Car: I assume E6.
User: No, E45.
Car: Showing traffic on the E45
```

If the user rejects the system suggestion without giving another answer, the system will show a menu where the most probable choices are the topmost ones, and marked in a distinct colour (not shown here).

```
User: Traffic information
Car: I assume E6.
User: No.
Car: What road?
Car GUI: [E6] [E45] [E20] [155]
```

When the system is certain about its hypothesis, the system will simply provide the user with the desired information without asking the user for parameters.

```
User: Traffic information
Car: Showing traffic on the E6
```

5 Conclusions and further work

We have designed and implemented a mechanism which learns user patterns and uses them proactively to simplify and shorten dialogue interactions. The idea of learning user patterns from observations is similar to Google Now. However,

while Google Now uses “cards” to provide unrequested information to the user, we show how predictions can be integrated into spoken or multimodal dialogue.

It remains for future work to evaluate the system to establish that this actually reduces the distraction rate of drivers. We also want to test the performance of the learning mechanism by training it on real observations of user behaviours (as opposed to simulated data).

The current mechanism only predicts answers to individual system questions, which may result in suboptimal behaviour in cases where there are dependencies between the questions pertaining to some task. An interesting area for future work is to instead predict sequences of answers; however, this would require a more powerful learning and classification mechanisms.

Acknowledgements

This work was carried out within the FFI project “Safe Speech by Knowledge” (2012-00941), funded by VINNOVA, Volvo Car Corporation and Talkamatic.

References

- H. H. Clark and S. E. Brennan. 1990. Grounding in communication. In L. B. Resnick, J. Levine, and S. D. Behrend, editors, *Perspectives on Socially Shared Cognition*, pages 127 – 149. APA.
- Dagmar Kern and Albrecht Schmidt. 2009. Design space for driver-based automotive user interfaces. In *Proceedings of the 1st International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, AutomotiveUI '09, pages 3–10, New York, NY, USA. ACM.
- Staffan Larsson, Alexander Berman, and Jessica Villing. 2011. Adding a speech cursor to a multimodal dialogue system. In *INTERSPEECH 2011, 12th Annual Conference of the International Speech Communication Association, Florence, Italy, 2011*, pages 3319–3320.
- Staffan Larsson. 2002. *Issue-based Dialogue Management*. Ph.D. thesis, Göteborg University.
- Tom M. Mitchell. 1997. *Machine Learning*. McGraw-Hill, New York.
- Vicki L. Neale, Thomas A. Dingus, Sheila G. Klauer, Jeremy Sudweeks, and Michael Goodman. 2005. An overview of the 100-car naturalistic study and findings. Technical report.